DOCUMENT RESUME

ED 093 911 TM 003 727

AUTHOR Molina, Huberto; Shoemaker, David M.
TITLE A Preliminary Evaluation of a Bilingual Spanish/English Program Using Multiple Matrix Sampling.
PUB DATE [73]
NOTE 35p.; Paper presented at the International Seminar on Language Testing (San Juan, Puerto Rico, May 1973)

EDRS PRICE MF-$0.75 HC-$1.85 PLUS POSTAGE
DESCRIPTORS Bilingual Education; Elementary School Students; *English (Second Language); Item Sampling; *Language Instruction; *Program Evaluation; Program Improvement; *Spanish Speaking
IDENTIFIERS Michigan Oral Language Series; MOL

ABSTRACT

The Michigan Oral Language program was installed in the Santa Paula School District for a 2-month tryout and student performance before and after the program was assessed through multiple matrix sampling. Student performance over seven content areas was determined for both the English and Spanish language components. In all seven content areas, performance was high in English and low in Spanish. Specifically, for each of the seven content areas investigated (mathematics-science, social studies, colors, numbers, commands, one-word answers, and total), performance in Spanish was approximately 58 percent less than that over the comparable English content area. The results suggest strongly that a major modification is required in the Spanish language component of the program. From exprience gained in this tryout, recommendations were made for improving the program. (Author/BB)

ED 093911 TM 003 227

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A PRELIMINARY EVALUATION OF A BILINGUAL SPANISH/ENGLISH PROGRAM USING MULTIPLE MATRIX SAMPLING

Huberto Molina

David M. Shoemaker

Paper presented at the International Seminar on Language Testing, jointly sponsored by TESOL (Teachers of English to Speakers of Other Languages) and AILA (International Association of Applied Linguistics) Commission on Language Tests and Testing, May 11, 1973, San Juan, Puerto Rico.

A PRELIMINARY EVALUATION OF A BILINGUAL SPANISH/ENGLISH PROGRAM USING MULTIPLE MATRIX SAMPLING

HUBERTO MOLINA

DAVID M. SHOEMAKER

Consultants To Santa Paula School District
Santa Paula, California

INTRODUCTION

Of the many variables which could be considered in the evaluation of a bilingual instructional program, some are more critical than others. We believe that the three most important variables are (a) the characteristics of the learners, (b) the learning environment, and (c) the linguistic components of the two languages.

The Learners

The age and entry skills of the students play an important role in any instructional program. In a bilingual program, they are even more important. The degree of maturation defines the linguistic structures which students can be expected to produce and determines reasonable criterion levels to be attained.

The Learning Environment

The learning environment of students must be viewed broadly. Fishman and Lovas (1970) suggest that program developers make their assumptions explicit about the kind of language situation which exists in a given community and to test the validity of these assumptions by gathering various kinds of data regarding the societal functions of community languages and existing attitudes towards them, both before and during the development of bilingual educacion programs. Tucker and d'Anglejan (1971) recommend a small scale survey of the local community in which the program will be situated. They recommend an assessment of the parents' language background and aspirations regarding their children's educational and occupational future. The bilingual competency of teachers and teachers' aides and their attitudes during the program are of no minor importance. Additional considerations are the manner in which the program is implemented and the class size.

Linguistic Components Of Languages

One of the primary concerns of most bilingual programs is having students acquire the linguistic skills necessary for communication in both languages. To account for recent research in the area of language acquisition, the linguistic structures of both languages are considered

within the generative transformational framework. In this framework, syntactic skills are stated in terms of phrase structure descriptions which underlie the basic sentence patterns and the transformations which these sentences may undergo. For example, a basic sentence pattern may be modified by (a) expanding elements, (b) by deleting elements, and by (c) permuting elements.

Expanding Elements	I see a book/I see a book and a pencil.
	Veo un libro/Veo un libro y un lápiz.
Deleting Elements	(The book is) on the table.
	(El libro está) en la mesa.
Permuting Elements	The book is on the table. On the table is the book.
	El libro está en la mesa. En la mesa está el libro.

Morphological variables account for changes that words undergo (e.g., want/wants, quiero/quiere) and phonological skills can be stated in terms of sound segments which students are required to produce in both languages.

The learner, the learning environment and the linguistic components of the languages are important factors in the evaluation of any bilingual program. However, due to time and cost considerations, the evaluation of the Santa Paula Bilingual Project was limited to the assessment of student competency over the linguistic components contained within both languages.

PURPOSE OF STUDY

The English/Spanish bilingual program considered herein was installed for two months in the Santa Paula School District in the Spring of 1972 on a trial basis. The thrust of the investigation was that of developing a procedure by which student performance over the linguistic components in both languages could be assessed in a rigorous and comprehensive manner. The purpose of the investigation was a determination of the strengths and weaknesses of the instructional program so that specific recommendations could be made concerning how best to expand the program into a full academic year.

DESCRIPTION OF PROGRAM

The Michigan Oral Language (MOL) Series -- Standard English as a second language -- was used as the basis for instruction. The methods used to implement the program were those specified by the authors of the program. The materials used within the program consisted of 59 English lessons and 61 Spanish lessons in addition to a Bilingual Conceptual Development Guide. The Spanish lessons are designed to teach Spanish-speaking students concepts first in their native language and are not designed for monolingual English-speaking students learning Spanish. The program assumes that Spanish-speaking teachers and aides are available. As described by the authors, the program is an ordered sequence of suggestions to teachers and they anticipate that the lessons will be used in many different ways depending upon the past experience of each teacher and the experience of the students with whom she is working. The authors emphasize the flexibility of their program by saying that many teachers will choose to utilize the lessons only at those times when they perceive a specific need for a certain type of oral language experience on the part of the students.

METHOD OF TESTING LINGUISTIC COMPONENTS

The primary emphasis in this investigation was an evaluation of the degree of mastery of criterion items and was not that of determining the relative achievement of individual students within the program. Although several evaluation procedures were possible, the one selected and considered most appropriate for the problem at hand was multiple matrix sampling. The rationale for selecting this procedure is stated simply. In order to evaluate comprehensively student performance in the MOL program, a large item population was required whose items span the entire range of objectives contained within the program. Of chief concern was how all students participating in the program would perform on all the items in the item population. Although it was not possible to administer all items to all students, it was possible to estimate (through multiple matrix sampling) the results which would have been obtained had this been done.

Constructing The Item Population

Each lesson in the instructional program was examined carefully and specific outcomes contained within that lesson were determined. Using these outcomes as the basis for constructing test items, 600 items were constructed -- 302 in Spanish and 298 in English. The Spanish items were generally a direct translation of the English items. Within each language, the items were subdivided into two large content strata: mathematics-science items and social studies items. Because the presentation of syntax within the program was graded from simple to difficult, the items within these two large strata were further subdivided according to difficulty level.

Multiple Matrix Sampling

In multiple matrix sampling (or, synonymously, item-examinee sampling), the item population is divided through random sampling into item subsets or subtests with each subtest administered to a subgroup of students selected randomly from those participating in the instructional program. In this procedure, different students respond to different subtests but the results may be used to estimate how all students would have performed over all the items in the item population. The advantages of multiple matrix sampling over alternative procedures for evaluating instructional programs are discussed in detail by Shoemaker (1972). The chief advantages of the procedure are the following:

1. A more comprehensive and detailed assessment is possible because it is student achievement on an item population -- instead of, for example, a standardized test or some item subset -- which is being determined. Were any item aggregate used in place of this item population, it is quite possible that the results would be specific to that particular test.

2. Because different students respond to different subtests (and not all students to one large item aggregate), a sharp decrease in testing time over traditional testing procedures is afforded.

3. In multiple matrix sampling, it is not necessary for all students participating in an instructional program to be tested. Both items and students are sampled in this procedure.

4. Multiple matrix sampling is suited ideally to pretesting and posttesting of student performance in an instructional program because subtests are assigned randomly to students on both occasions. This avoids the problems (e.g., the carry-over) which may occur when the same test is administered on both occasions.

5. Estimates of student performance obtained from multiple matrix sampling are more accurate on the average than those obtained from any item subset.

Construction Of Subtests

In this investigation, the item population consisted of 600 items subdivided into four major content strata: (a) 200 Spanish language items dealing with mathematics and science, (b) 103 Spanish language items dealing with social studies, (c) 199 English language items dealing with mathematics and science, and (d) 99 English language items dealing with

social studies. Items within each stratum were subdivided further according to the syntactic structure of both the question and the minimum acceptable answer into three levels of difficulty. The item population was then subdivided into 20 subtests containing 30 items[1] each by stratified-random sampling (i.e., items were sampled randomly from the population such that the proportion of items in the test from each of the 12 strata was approximately the same as that found in the item population.) The first of the 20 subtests is given in the Appendix. The allocation of items to all subtests is given in Table 1.

Student Population

Ten elementary grade classes participated in the tryout of the MOL program: 7 kindergarten classes, 2 first grades, and one combination first and second grade. The total number of students involved was approximately 300.

Test Administration Procedures

Fifteen students selected randomly from within each of the 10 classes were pretested at the beginning of the program and posttested at the end. At both testings, students were selected randomly from each class. The time interval between the two testings was approximately two months.

Both pretesting and posttesting were accomplished similarly. Five subtests, selected randomly from the battery of 20, were assigned to each class and each subtest administered to three students. Students were selected randomly from each class but subject to the restriction that, over classes, each subtest was administered to approximately the same number of males and females. Additionally, when selecting students for the posttesting, if a particular student selected had also been pretested, he was always given a subtest different from the one taken originally. Over classes, the number of students responding to each subtest was, generally, either 6 or 9. The total number of students pretested was 146 and, posttested, 148. In assigning subtests to classes, it should be noted that several plans could have been used. For example, one alternative procedure would have been to randomly assign two subtests to each class with from 7 to 8 students responding to each subtest. In this procedure, the ability of each class would have been reflected by 60 items and, what is more important, their performance could have been a function of the specific items included in the subtests. In place of this, five subtests totalling 150 items were used within each class to minimize this problem and to expose each class to a broad item base. In both cases, however, approximately the same number of students respond to each subtest.

---

[1] All subtests were designed originally to contain 30 items. Subtest 17, however, contained only 29 items due to a typographical error.

Table 1

Number And Content Reference Of Items Allocated To Subtests

Subtest	Number of Items	Spanish		English	
		MS	SS	MS	SS
01	30	11	5	9	5
02	30	10	6	9	5
03	30	11	5	10	4
04	30	9	5	11	5
05	30	10	5	10	5
06	30	9	6	10	5
07	30	11	5	9	5
08	30	9	5	11	5
09	30	11	5	10	4
10	30	9	5	11	5
11	30	11	5	9	5
12	30	9	5	11	5
13	30	11	5	9	5
14	30	9	5	11	5
15	30	10	5	9	6
16	30	9	6	10	5
17	29	10	5	10	4
18	30	10	5	10	5
19	30	10	5	10	5
20	30	10	5	10	5
Total	599	199	103	199	98

MS = Mathematics/Science Items
SS = Social Studies Items

Each subtest was administered individually to each student by a bilingual female tester hired specifically for the task. Before pretesting and posttesting, all 5 testers used in this investigation were given an orientation and training program. The average testing time per student was approximately 6 minutes. If an individual student answered incorrectly five consecutive items in either language section, the remaining items in that section were omitted and scored as incorrect. All items were scored dichotomously (1 = pass, 0 = fail).

Parameters Estimated

The statistical procedures for estimating parameters through multiple matrix sampling used in this investigation were those described by Shoemaker (1971a). In general, the mean test score and variance of test scores over the four major content strata were estimated at the beginning and end of the instructional program. Using estimates of these parameters and the curve fitting procedures described by Shoemaker, normative distributions were estimated for all four content strata both pretest and posttest.

RESULTS

Estimates of the mean number correct test score for all content strata considered in this investigation are given in Table 3 for the Spanish language items and in Table 4 for the English language items. (At this point, the reader should ignore the 'Color Items', 'Number Items', etc. results.) The 'All Items' stratum is merely the pooling of the 'Mathematics-Science Items' with the 'Social Studies Items.' The estimated normative distributions for the four major content strata and the two combined strata for both pretest and posttest are plotted in Figures 1 through 6. Because the graphs in each Figure are interpreted similarly, only those in Figure 1 will be described in detail.

In Figure 1 are plotted two normative distributions -- one for the pretest scores 'Pre' and one for the posttest scores 'Post' -- for the 302 Spanish language items in the item population. Each curves gives the estimated relative frequency distribution which would have been obtained had all 300 students been tested individually over the 302 items comprising this content stratum. The height of the curve above each test score gives the proportion of students expected to receive that particular test score (number of items answered correctly). For example, .0130 (or, 1.30 per cent) of the students would be expected to answer exactly 25 items correctly on the 302-item pretest and .0126 (or, 1.26 per cent) would be expected to receive the same score when the test was administered at the end of the program.

For the six content strata considered in Figures 1 through 6, the estimated pretest performance of the students did not differ significantly from that of their posttest performance. In this investigation, the standard

Table 2

Reclassification Of Items Within Subtests

Subtest	Spanish				English			
	Color	Number	Command	One-Word	Color	Number	Command	One-Word
01	2	2	1	4	0	2	1	4
02	3	1	2	4	3	2	1	5
03	1	0	3	6	3	2	1	5
04	1	1	1	10	0	0	1	6
05	0	1	1	6	0	1	1	3
06	2	0	3	4	1	0	1	6
07	3	1	2	2	2	2	1	3
08	0	1	1	3	0	0	2	2
09	0	1	1	8	2	1	2	4
10	2	1	3	5	2	0	6	6
11	2	1	1	6	1	2	0	5
12	1	0	3	0	1	0	2	4
13	1	3	1	9	0	1	2	2
14	0	3	1	7	0	0	3	2
15	2	1	0	5	1	1	2	6
16	0	0	4	4	3	0	0	6
17	1	2	2	4	1	2	1	7
18	1	2	3	5	4	3	1	10
19	2	1	0	4	3	2	3	6
20	1	2	2	7	1	1	1	5

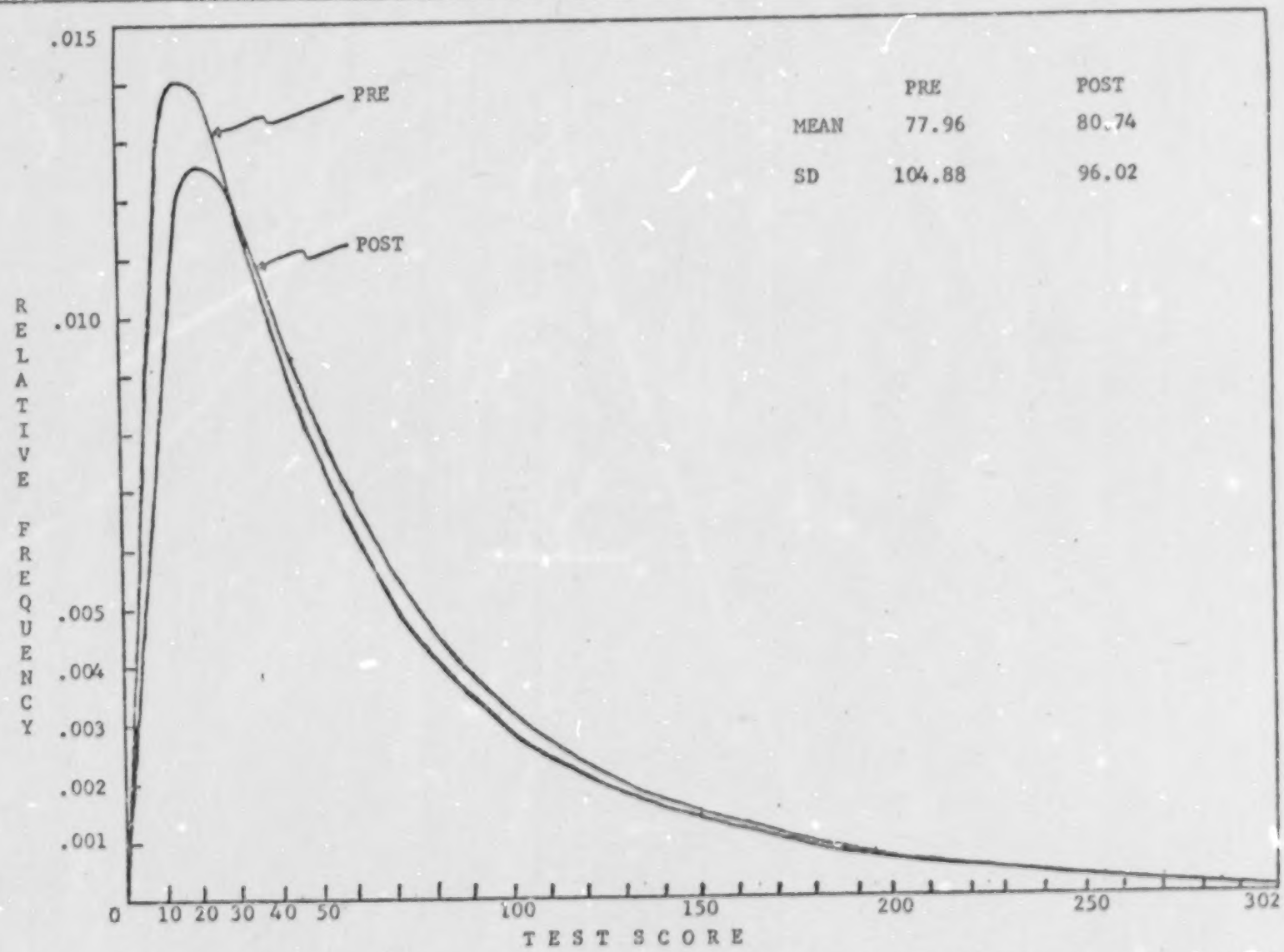


Figure 1: Estimated Pretest And Posttest Normative Distributions For All Spanish Language Items

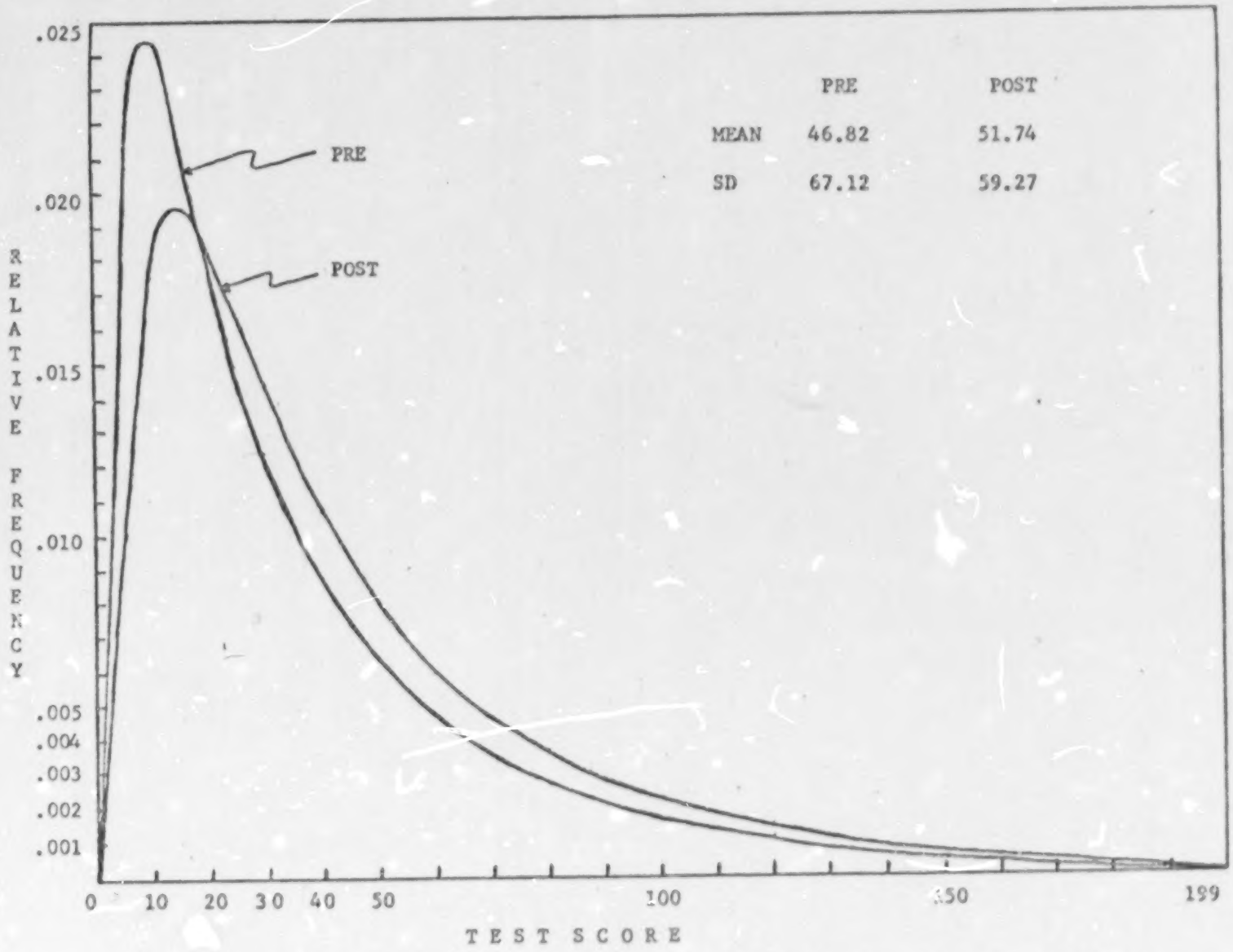


Figure 2: Estimated Pretest And Posttest Normative Distributions For Spanish Language Mathematics/Science Items

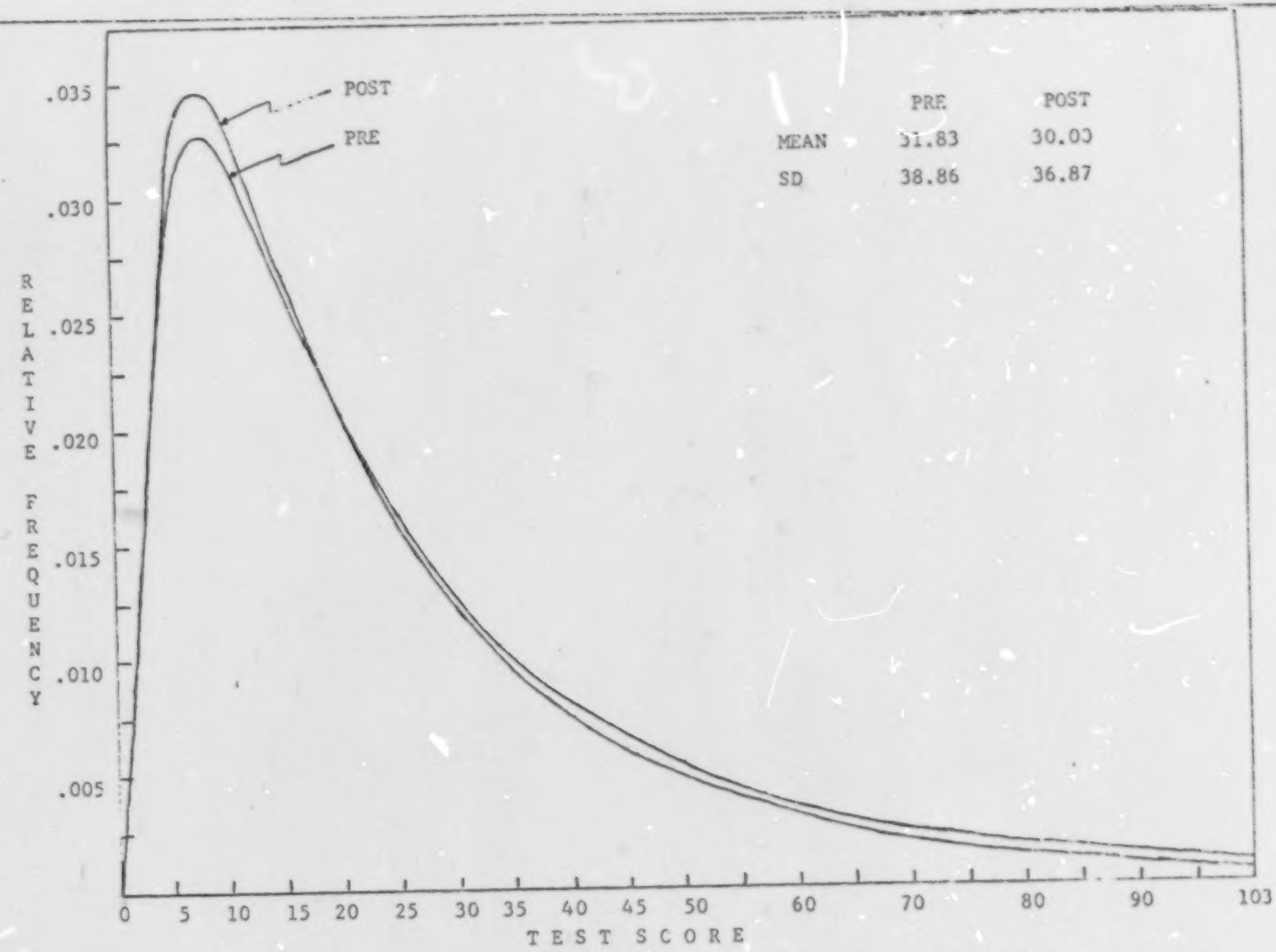


Figure 3: Estimated Pretest And Posttest Normative Distributions For Spanish Language Social Studies Items

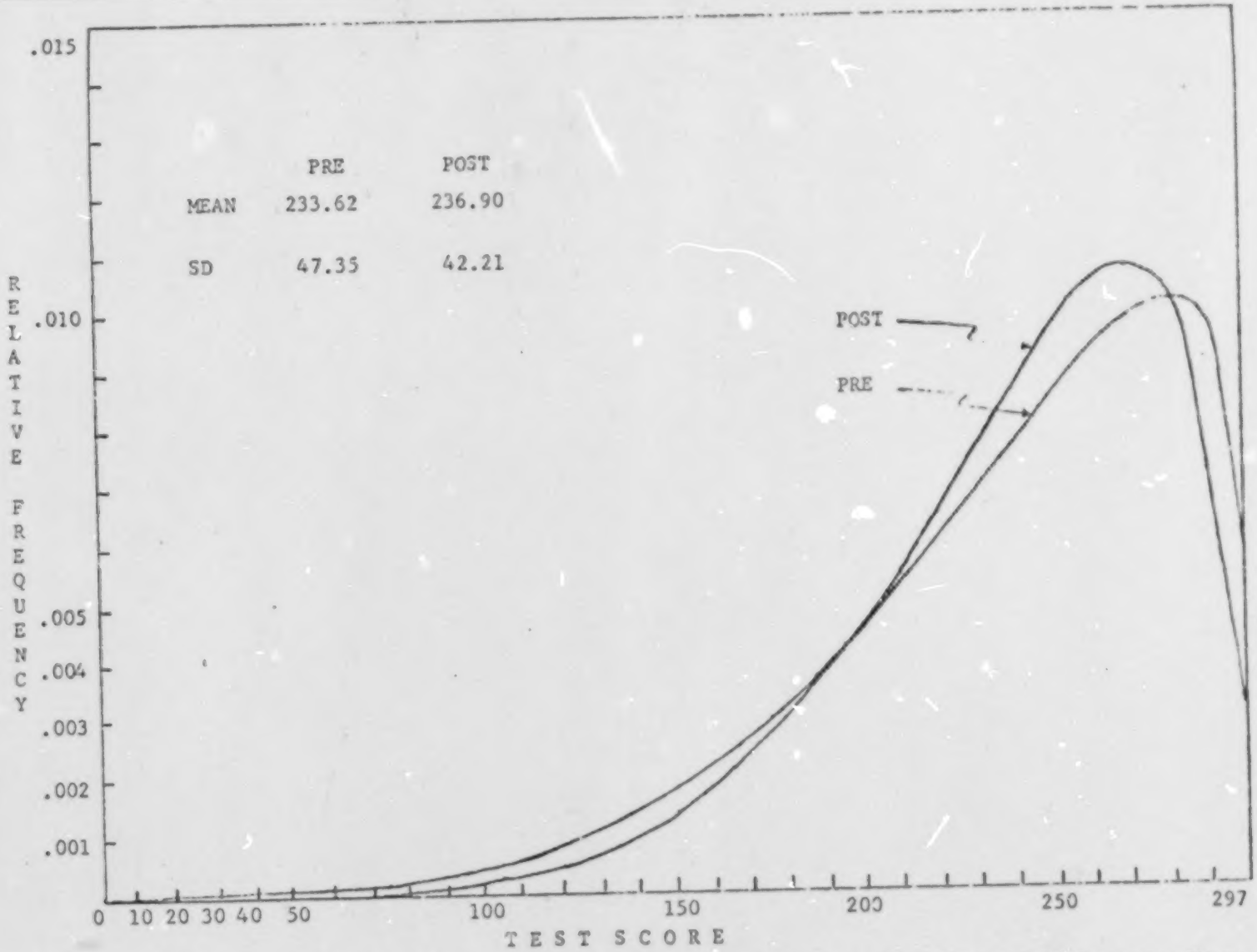


Figure 4: Estimated Pretest And Posttest Normative Distributions For All English Language Items

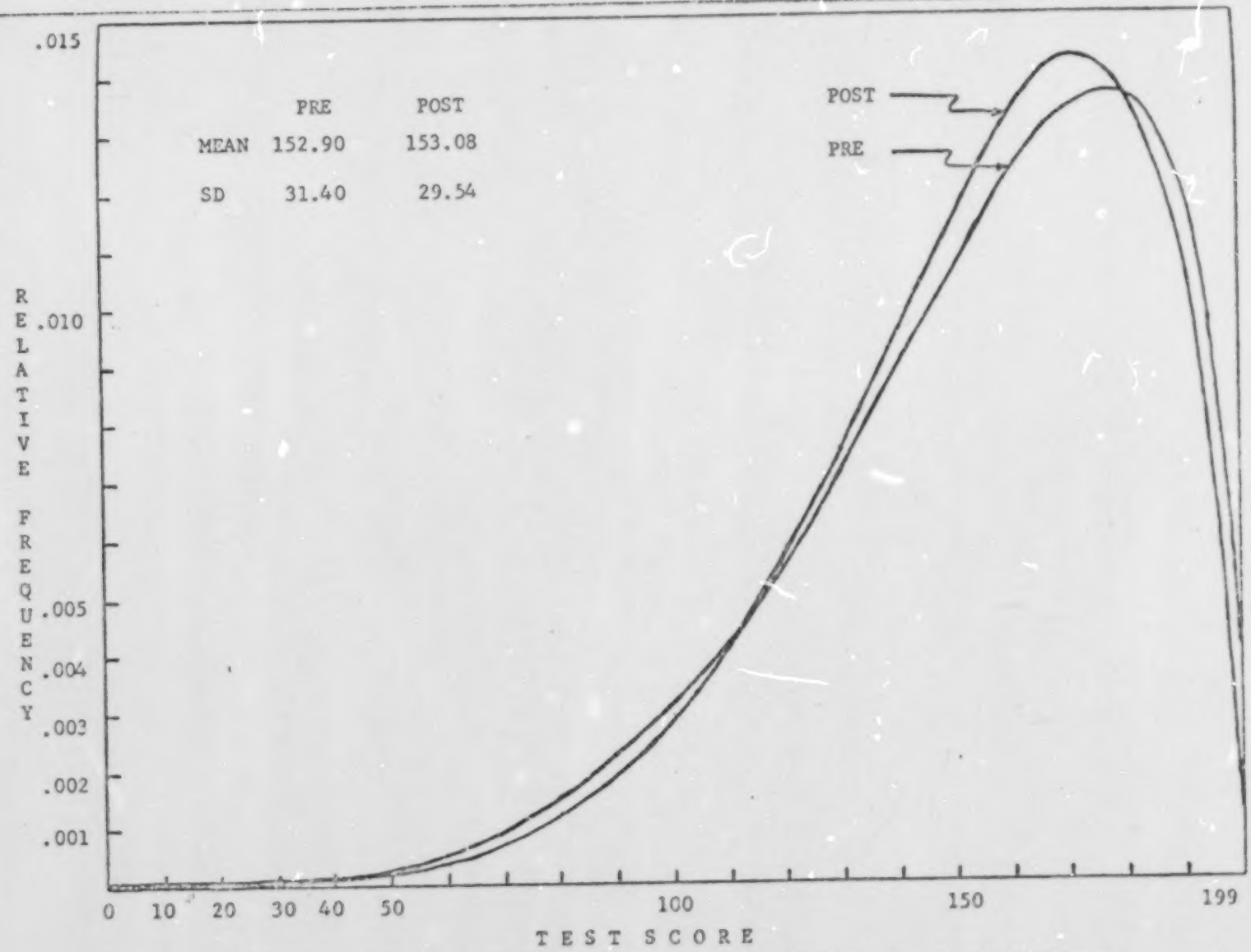


Figure 5: Estimated Pretest And Posttest Normative Distributions For English Language Mathematics/Science Items

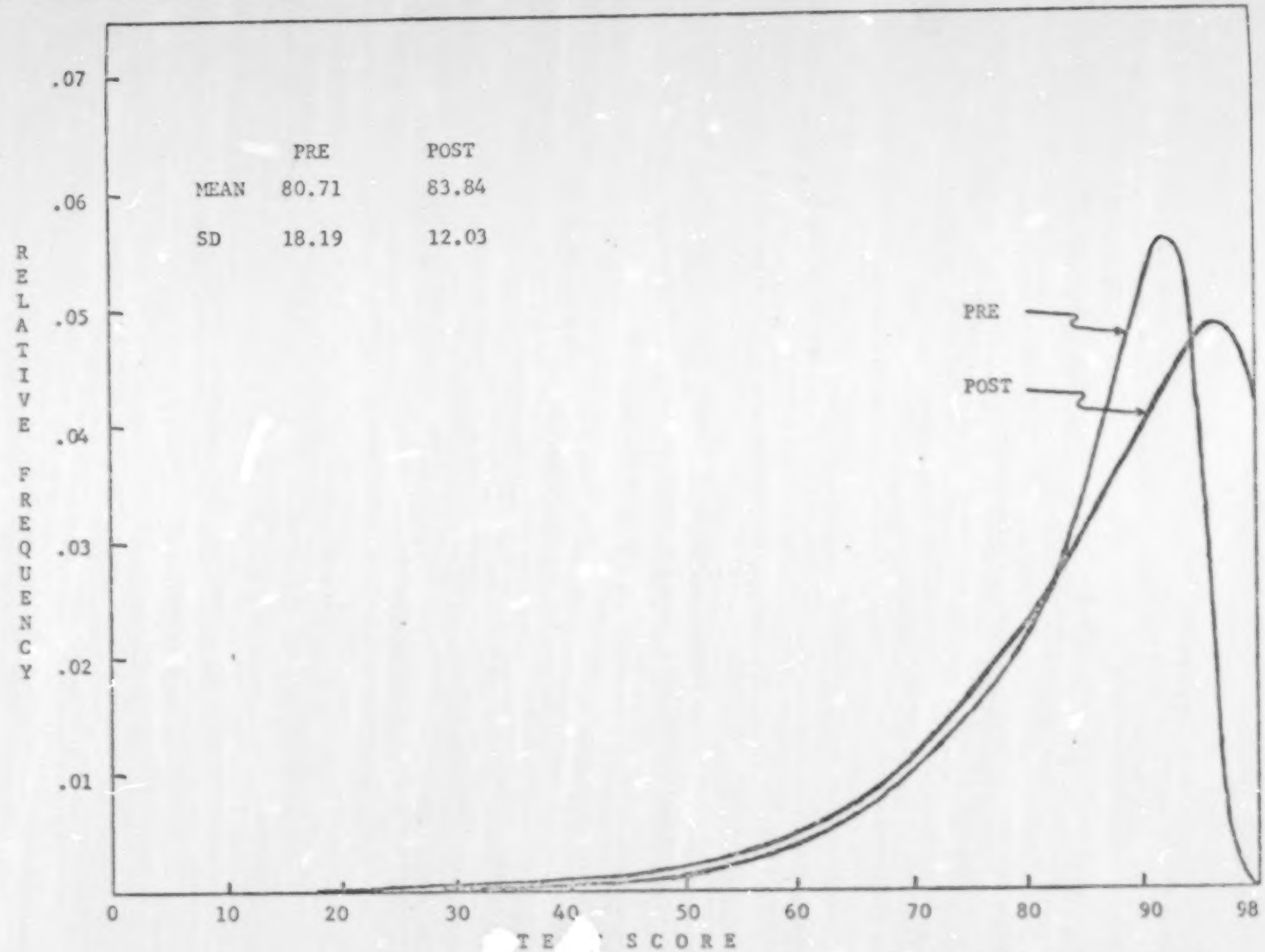


Figure 6: Estimated Pretest And Posttest Normative Distributions For English Language Social Studies Items

Table 3

Estimated Mean Scores For Selected Content Strata At Pretest And Posttest For Spanish Language Items

Content Strata	Number of Items	Pretest	Posttest
All Items	302	77.96 (.258)*	80.74 (.267)
Mathematics/Science	199	46.82 (.235)	51.74 (.260)
Social Studies	103	31.83 (.309)	30.00 (.291)
Color	25	4.98 (.199)	5.59 (.224)
Number	24	7.54 (.314)	7.18 (.299)
Command	35	9.88 (.282)	9.95 (.284)
One-Word Answer	103	30.21 (.293)	27.89 (.271)

*Proportion correct: 77.96/302 = .258

Table 4

Estimated Mean Scores For Selected Content Strata At Pretest And Posttest For English Language Items

Content Strata	Number of Items	Pretest	Posttest
All Items	297	233.62 (.787)*	236.90 (.798)
Mathematics/Science	199	152.90 (.768)	153.08 (.769)
Social Studies	98	80.71 (.824)	83.84 (.856)
Color	28	25.27 (.903)	23.89 (.853)
Number	22	18.04 (.820)	20.87 (.949)
Command	32	28.56 (.893)	28.26 (.883)
One-Word Answer	97	79.30 (.816)	83.64 (.862)

*Proportion correct: 233.62/297 = .787

error of estimate for each parameter was approximated conservatively by the jackknife procedure and, on the basis of these estimated standard errors, the estimates of parameters for pretest and posttest scores were judged to be not significantly different. Although the differences were not significant, it should be noted that the estimated variance of the posttest scores was always less than that of the pretest scores.

The lack of significant pretest-posttest differences for each of the six content strata considered in Figures 1 through 6 suggested that the effects of the MOL program were at best slight and possibly found only in smaller segments of the item population. For this reason, items within each subtest were reclassified according to (a) color -- items dealing with colors, (b) number -- items dealing with number concepts, (c) command -- items requiring the student to carry out a command, and (d) one word answer -- items having a minimal acceptable answer of one word. These four content strata were selected because basic colors, numbers 1 through 5, commands requiring no verbal response, and one-word responses represent very elementary skills in both subject matter and linguistic performance. The total number of items in each of these four content for both languages per subtest are given in Table 2. Estimated mean test scores for these eight additional content strata are given in Tables 3 and 4. Once again, the pretest results were judged not to differ significantly from posttest results. Estimates of the test score variance and other parameters were not considered with these additional content strata because of the small number of items within each subtest on which they would be determined.

DISCUSSION AND RECOMMENDATIONS

The Michigan Oral Language program was presented during the final two months of the school year and, for this reason, it was quite possible that students entering the program had already acquired many of the fundamental linguistic skills needed to communicate in the classroom. The high pretest mean score on all items in the English language stratum suggests that this was, indeed, the case with English. Had the pretest been presented at the beginning of the school year, greater gains in English might have resulted. Such was not the case with the estimated pretest mean score of all students on the Spanish language stratum. Here, student achievement was quite low -- both pretest and posttest. This result suggests that the Spanish language component of the program will require a major modification to bring student performance on Spanish up to par with that of English. It should be noted that even monolingual Spanish-speaking students scored poorly on individual subtests. The reason for this is possibly that the Spanish vocabulary items in the program dealt with content areas discussed infrequently by these students within their home environment.

This investigation was successful in establishing a procedure for testing linguistic components in an efficient and inexpensive manner and the results obtained indicate clearly the relative emphasis in each content area needed to bring about bilingual performance. We suggest that instruction

in the MOL be modified in the following ways:

1. Materials

New materials should be developed or the MOL materials should be revised and implemented with additional materials to provide prerequisite language skills for using both Spanish and English as media of instruction. Were this to be done, the materials would be consistent with Andersson's (1968, 1970) definition of bilingual education. The choice -- either to revise the MOL or to develop new materials -- is based on the availability of funds. In either case, the following should be done:

a. The first structures of both languages introduced should be the most basic and the progression in the presentation of syntax and concepts should lead the learner to sentences of greater complexity. The students should be equipped with basic language competencies and strategies by which to add to linguistic competency. Language skills should include syntax, phonemic control, and intonation. These linguistic requirements and lexicon are discussed in greater detail by Molina (1971a). The program should be tied to the tested entry skills of the students.

b. Objects used in the program should be easily accessible to the teacher and selected to make conceptual distinctions explicit. A ball is an example of one object which may be used advantageously.

English		Spanish	
a ball		una pelota	
It's a ball.	(to be)	Es una pelota.	(ser)
There is a ball on the table.	(to be)	Hay una pelota en la mesa.	(haber)
The ball is on the table.	(to be)	La pelota está en la mesa.	(estar)

Through the manipulation of common objects the linguistic patterns of both languages are revealed. By the use of a ball/pelota, the three Spanish forms 'ser/haber/estar' are introduced in a meaningful way. It has been suggested by Molina (1970) that the presentation of classroom items must be organized in such a way that the learner becomes aware that he is not learning isolated words but is, instead, learning sets of words which operate in a systematic way.

c. Tapes should be developed for the teachers as an aid in presenting each Spanish lesson. It should not be assumed that the teachers are bilingual. Even among those teachers who are bilingual, degrees of bilingualism are to be expected. Tapes would serve to standardize the Spanish lesson across teachers having differing degrees of competency in Spanish. Additionally, it cannot be assumed that all teachers will have the assistance of Spanish-speaking aides.

d. The format for each lesson should include a list of objectives stated in behavioral terms. The lesson should state explicitly the learning activities that are to bring about change in student behavior. Each lesson should include two types of activities: teacher/learner and learner/learner. In the former, the teacher establishes communication with the student. She communicates with the entire class and with individual students. In the latter activity, students use newly acquired language skills to talk with one another. In both of these activities, the teacher is able to evaluate the effectiveness of her teaching in terms of student response. Each lesson should end with a play activity that relates directly to the objectives of the lesson. One example of such a play activity is 'Going To The Bank.'

Going To The Bank
A student cashes a check marked from 1 to 10 and receives a corresponding number of cut-out dollar bills. He counts each bill to be sure that he has received the correct number of bills.

It should be noted that play activities provide an excellent means of reviewing mathematics-science items.

e. Supplementary materials should be developed that can be presented by teachers, aides or tutors. The design of games that encourage verbal interaction between players and tutor with a minimum of adult supervision have been described by Molina (1971b). Tutorial materials should review the day's lesson and provide a cumulative review of previously learned material. Cumulative reviews should be structured such that students are asked to produce previously presented syntax in new conceptual contexts. Thus, students are encouraged to produce sentences beyond that which they have been taught.

2. Assessments Should Be Built Into The Program

   a. A Placement Test should be included as part of the program. The test should assess the competency of each student in both languages. The entry skills which should be tested are phonological control, syntax, conceptual understanding, and non-verbal responses (i.e., following directions.) A Placement Test given at the start of the program provides the basis for placing each student in the program and, as such, should enable the teacher to form language ability groups when feasible to do so. Guidelines for the construction of such tests are given by Valette (1967).

   b. Periodic assessments should be included at ten-lesson intervals which inform the teacher of needed remediation and provide a method by which she can present certain portions of a lesson at a faster pace. Each block assessment should be a trident-test -- a test having items on the most recent block, items from past blocks covering major outcomes and a sampling of items from the next instructional block of 10 lessons. A more detailed discussion of periodic assessments such as these is given by Shoemaker (1971b). Remediation exercises should be available for students whose performance is not acceptable.

   c. Teachers should assess the program on a monthly basis commenting on student, teacher and parent reaction to the program and any suggested program modifications. Comments such as these are solicited readily by a teacher questionnaire. An example of such a questionnaire easily completed by teachers is given by Molina (1972).

   d. Classroom observations should be scheduled periodically.

3. Pretesting And Posttesting Of All Program Objectives

   All criterion items in the program should be pretested and posttested using multiple matrix sampling.

SUMMARY

The Michigan Oral Language program was installed in the Santa Paula School District for a two-month tryout and student performance before and after the program was assessed through multiple matrix sampling. Student performance over seven content areas was determined for both the English and Spanish language components. In all seven content areas, performance was high in English and low in Spanish. Specifically, for each of the seven content areas investigated (mathematics-science, social studies, colors, numbers, commands, one-word answers, and total), performance in Spanish was approximately 58 per cent less than that over the comparable English content area. The results suggest strongly that a major modification is required in the Spanish language component of the program. From experience gained in this tryout, the following recommendations were made for improving the porgram:

1. Develop a new instructional program or reorganize and implement the MOL. Regardless of the choice made here:

   a. Provide explicit directions for the teacher as part of the format for each lesson.

   b. Provide tapes in Spanish.

   c. Provide materials for tutors or aides that relate directly to the objectives of the program.

2. Provide built-in evaluation procedures within the instructional program to inform the evaluator and teachers of student progress on a systematic basis using:

   a. A Placement Test

   b. Periodic assessments at 10-lesson blocks

   c. Remediation exercises

   d. Classroom observation forms

   e. Teacher questionnaires

3. Through multiple matrix sampling, determine the degree of mastery on all criterion items at the beginning and end of the program.

REFERENCES

Andersson, T. Bilingual elementary schooling: a report to Texas educators. Florida FL Reporter, 1968, 6, 3-4.

Andersson, T. & Boyer, M. Bilingual schooling in the United States. Washinton, D. C.: U. S. Government Printing Office, 1970.

Fishman, J. & Lovas, J. Bilingual education in sociolinguistic perspective. TESOL Quarterly, 1970, 3, 215-222.

Molina, H. Scientific and pedagogical grammars. Hispania, 1970, 53, 77.

Molina, H. The learner, the teacher, the grammar and the method in designing an instructional program. Hispania, 1971, 54, 439-444. (a)

Molina, H. Language, games and the Mexican-American child learning English. TESOL Quarterly, 1971, 5, 145-148. (b)

Molina, H. Evaluation during development of Spanish materials on the kindergarten and first grade level. Hispania, 1972 (In Press).

Shoemaker, D. M. Evaluating the effectiveness of competing instructional programs. Educational Researcher, 1972, 1, 5-8.

Shoemaker, D. M. Principles and procedures of multiple matrix sampling. Southwest Regional Laboratory for Educational Research and Development, Technical Report No. 34, 1971. (a)

Shoemaker, D. M. Criterion-referenced measurement revisited. Educational Technology, 1971, 11, 61-62. (b)

Tucker, G. R. & d'Anglejan, A. Some thoughts concerning bilingual education programs. The Modern Language Journal, 1971, 55, 491-493.

Valette, R. Modern language testing. New York: Harcourt, Brace & World, 1967.

APPENDIX

Subtest Number 1

Content Reference Of Each Item In Subtest Number 1

Item	Spanish						English					
	MS	SS	Cl	Nm	Cm	OW	MS	SS	Cl	Nm	Cm	OW
01	x											
02	x		x									
03	x											
04	x					x						
05	x			x								
06	x			x		x						
07	x											
08	x											
09	x											
10	x											
11	x					x						
12		x										
13		x			x							
14		x										
15		x				x						
16		x	x									
17							x					
18							x					
19							x					x
20							x			x		x
21							x			x		x
22							x					
23							x					
24							x					
25							x					
26								x				
27								x				
28								x				x
29								x				
30								x			x	

MS = mathematics-science item
SS = social studies item
Cl = color item
Nm = number item
Cm = command item
OW = one word answer item

BILINGUAL PROGRAM
SANTA PAULA SCHOOL DISTRICT

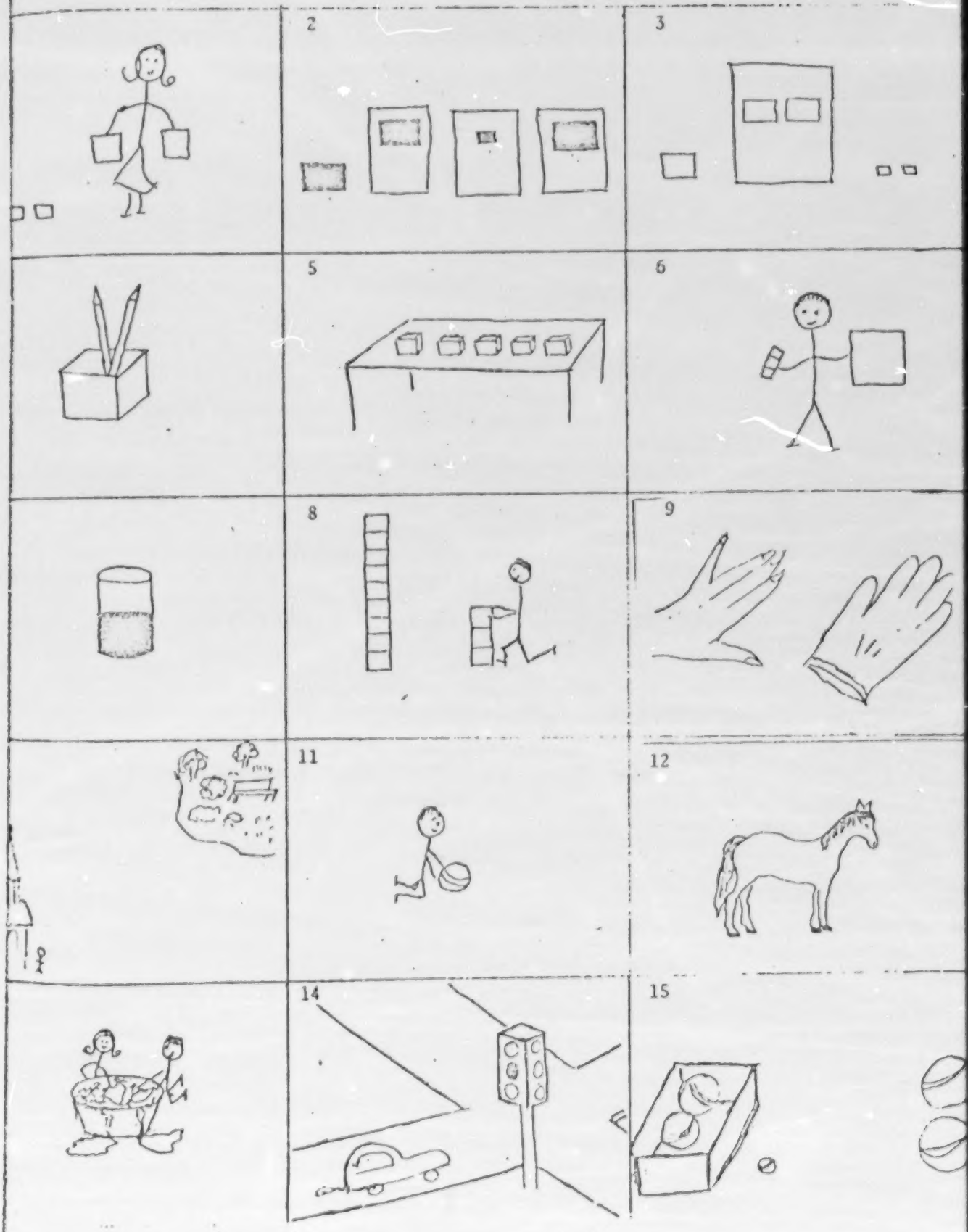
Directions are in brackets.
Tester Script is in capitals.
Acceptable answers are underlined.

Test No. 1

Materials Needed: 1 pencil, 1 crayon

Tester Script	Student Response. Shortest possible correct answer is underlined.
1. (Point to Picture 1) ¿CUÁLES BLOQUES TIENE ANA?	Los grandes.
2. (Point to object in Picture 2) ¿POR QUÉ VA ESE BLOQUE AQUÍ?	Porque es azul y grande.
3. (Point to object in Picture 3) ¿POR QUÉ VA ESTE BLOQUE EN ESA CAJA?	Porque es grande.
4. (Show 1 pencil and 1 crayon; take pencil away) ¿CUAL FALTA?	El lápiz.
5. (Point to objects in Picture 4) ¿CUÁNTOS LÁPICES HAY EN LA CAJA?	Hay dos lápices.
6. (Point to objects in Picture 5) ¿CUÁNTOS BLOQUES HAY EN LA MESA?	Hay cinco.
7. (Point to Picture 6) ¿DE CUÁLES TIENE MAS?	De los cuadrados pequeños.
8. (Point to Picture 7) ¿CÓMO ESTÁ ESE VASO?	Por la mitad.
9. (Point to Picture 8) PRIMERO HIZO UNA TORRE ALTA. ¿QUÉ HIZO DESPUES?	Una torre baja.
10. (Point to Picture 9) ¿PORQUÉ VAN JUNTOS?	Porque el guante se pone en la mano.
11. (Point to Picture 10) EL MUCHACHO ESTÁ CERCA DE IA IGLESIA. ¿ES CERCA DE LA IGLESIA AL PARQUE?	No, es lejos.
12. (Point to Picture 11) ¿EMPUJA O JALA LA BOLA?	Empuja la bola.
13. (Child has hands on table) BAJEMOS IA MANO.	(Child lowers hand to side.)
14. (Point to Picture 12) ¿QUÉ ES ESTE?	Es un caballo.
15. (Point to Picture 13) ¿QUÉ ESTÁN HACIENDO ELLOS? BRINCANDO O LAVANDO?	Están lavando.
16. (Point to Picture 14) ¿POR QUÉ PARÓ EL CARRO?	Porque la luz esta amarilla.

TEST 1 Tester Script	Student Response. Shortest possible correct answer is underlined.
17. (Point to Picture 15) WHY DO THESE BALLS GO IN THIS BOX?	Because they're big.
18. (Point to Picture 16) WHAT ARE THEY DOING? RUNNING SLOW TOWARD THE DOOR OR TOWARD THE CHAIR?	They're running slow toward the chair.
19. (Point to Picture 17) WHAT DID HE PUT OVER THE PAPER?	He put the triangle over the paper.
20. (Point to Picture 18) HOW MANY BOXES DOES HE HAVE?	He has one box.
21. (Give crayon to child) HOW MANY CRAYONS DO YOU HAVE?	I have one.
22. (Point to Picture 19) WHAT DID HE DRAW FIRST, A TALL OR A SHORT TOWER?	He drew a tall tower.
23. (Point to Picture 20) WHAT IS THAT?	It's a square.
24. (Point to Picture 21) WHERE DID HE PUT THE CUP, OVER OR UNDER THE PLATE?	He put the cup over the plate.
25. (Point to Picture 22) WHICH TRIANGLE DID HE PASTE FIRST?	The big one.
26. (Show Picture 23 and put away) DID HE PUSH OR PULL THE CAR?	He pushed the car.
27. (Point to Picture 24) WHAT IS IT? A BALL OR A BLOCK?	A block.
28. (Point to Picture 25) THIS IS ANA AND JOSÉ. WHO HAS THE HORSES?	José has them.
29. RAISE YOUR HAND.	(Child carries out command)
30. (Point to Picture 26) THIS IS ROSA; THIS IS LUIS. WHOSE FOOTPRINT IS THIS?	It's Rosa's.

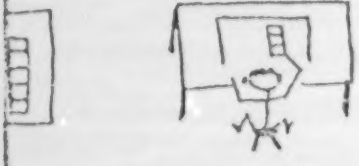

2
3
5
6
8
9
11
12
14
15

FICTURES

TEST 1

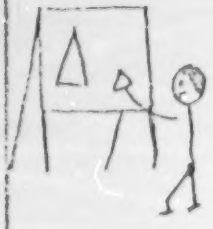
17

18



20

21



23

24

26